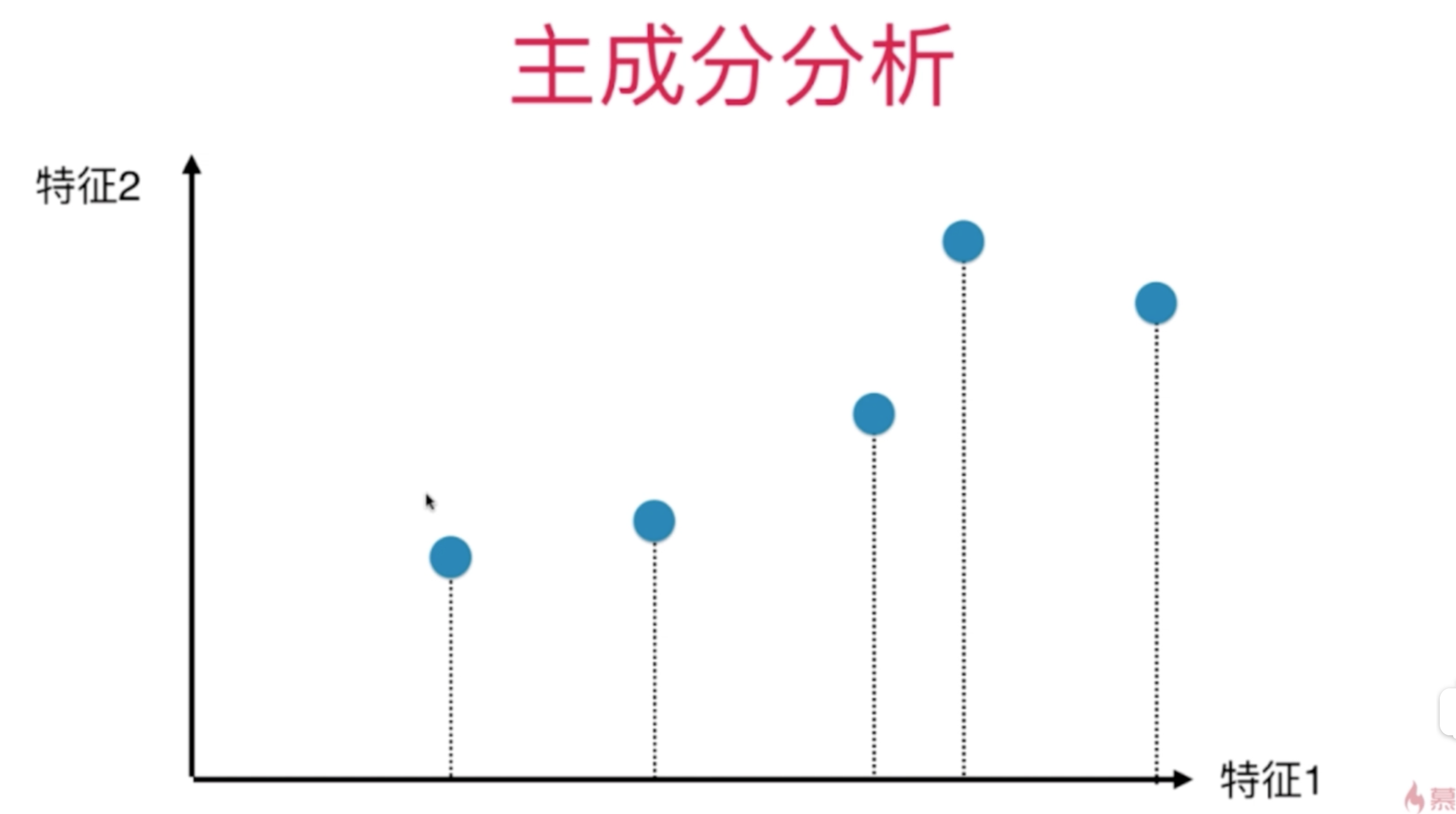
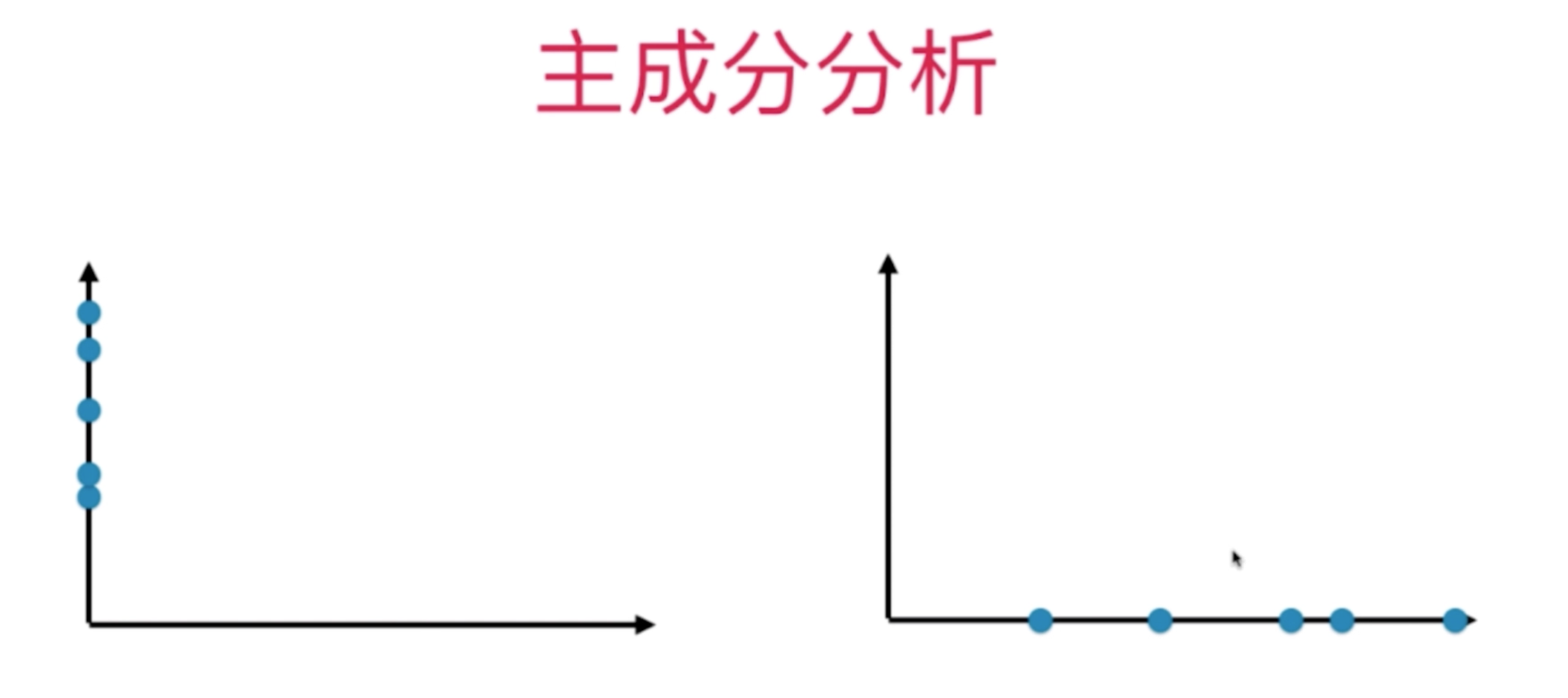


PCA 是统计学、数理统计学科的重要方法。



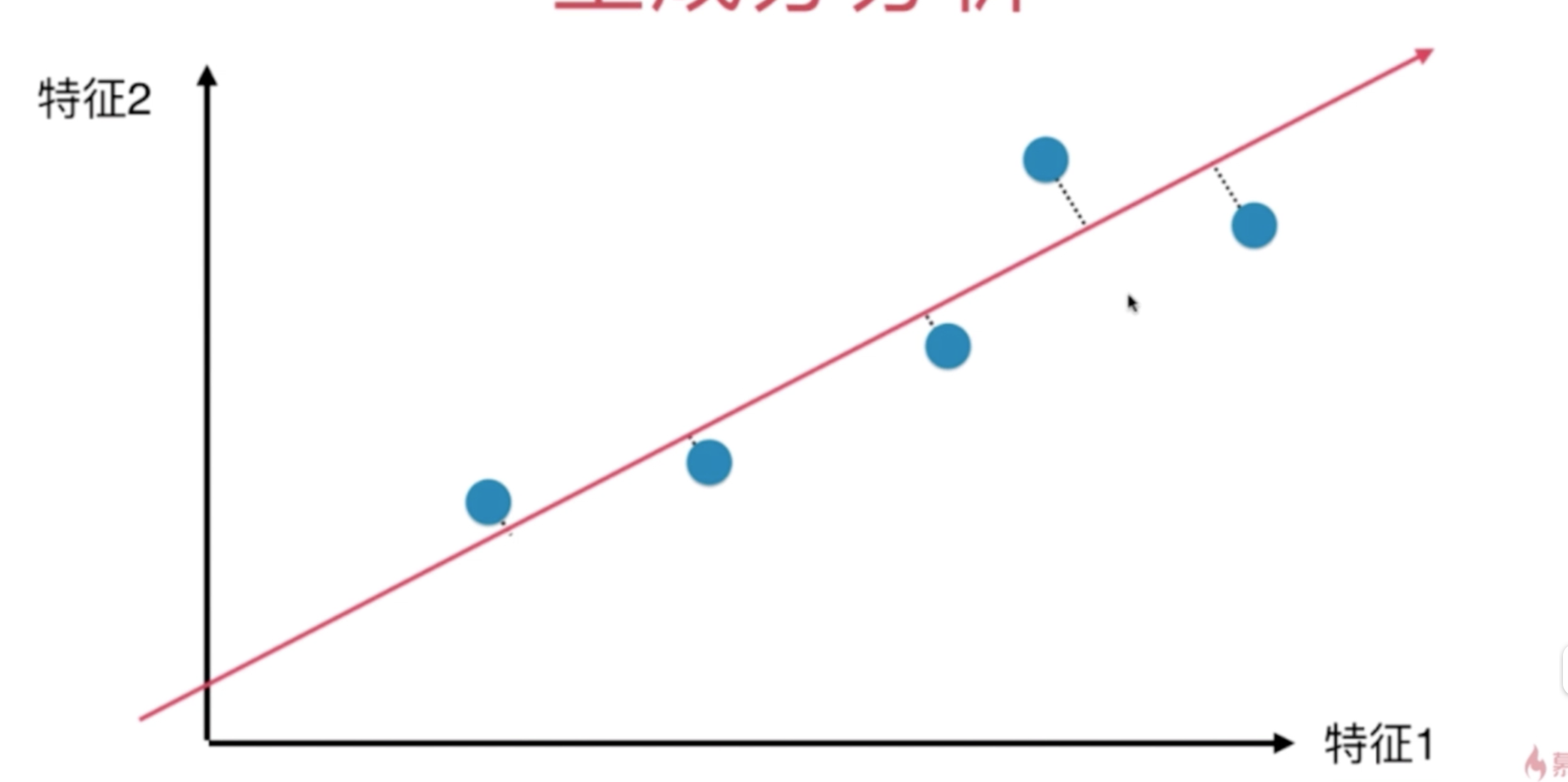


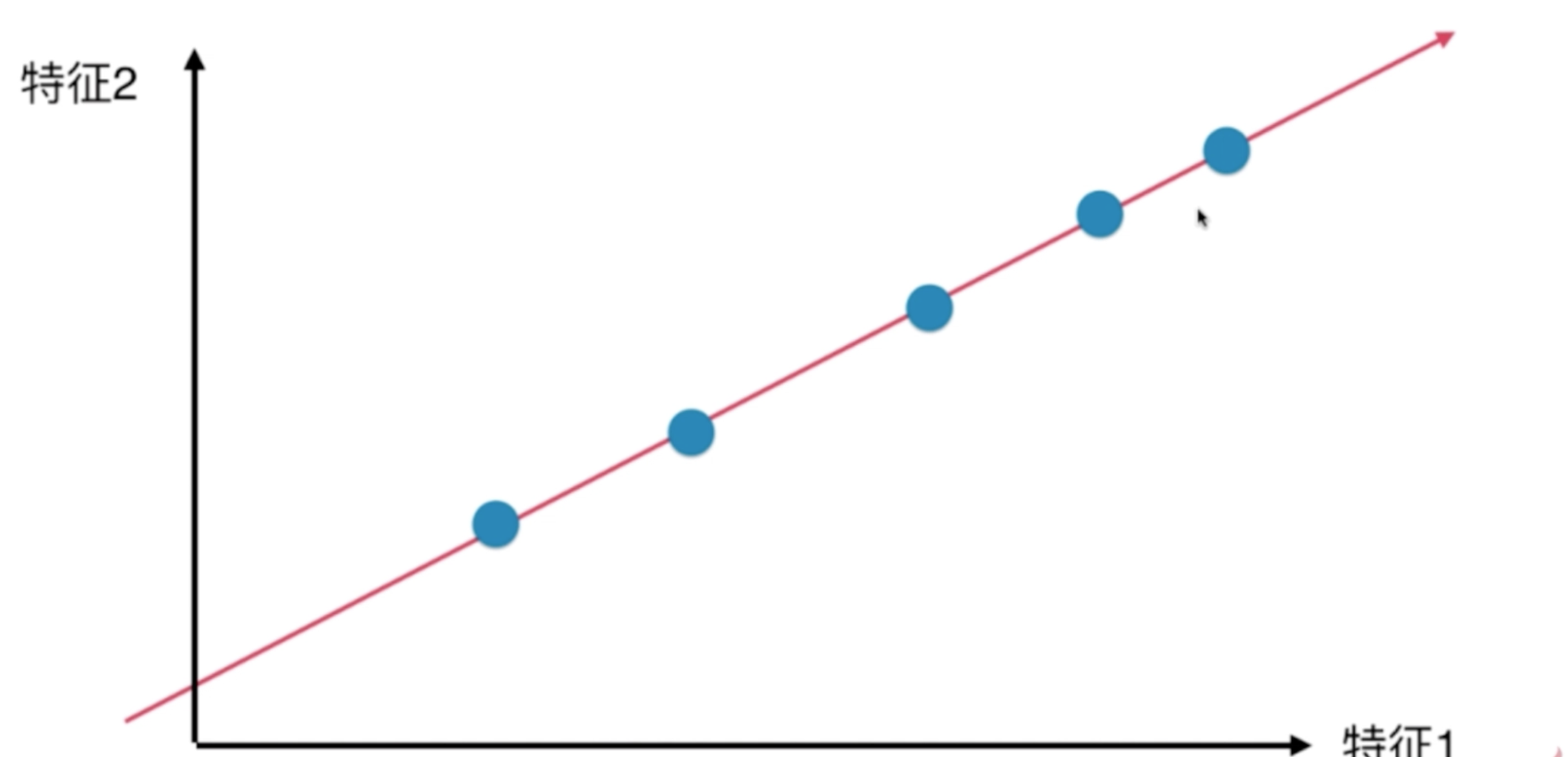
如何降维？最简单的就是把其中一个特征的数据扔掉



左边的是扔掉特征1，右边的是扔掉特征2，显然右边的是更好的降维方案，因为：点和点之间的距离时相对比较大的，更能体现样本原来的间距特征

但是这是最好的方案吗？

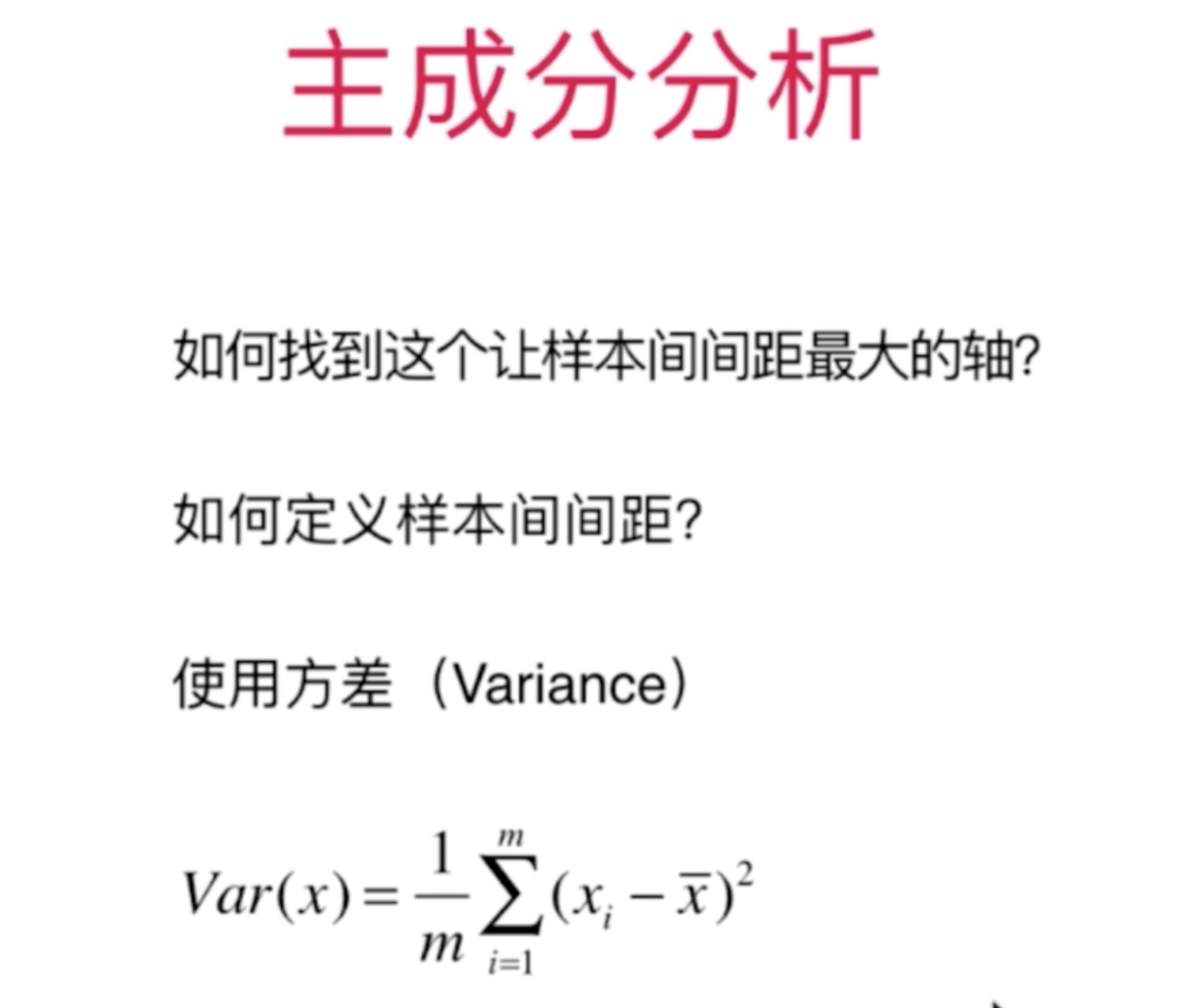


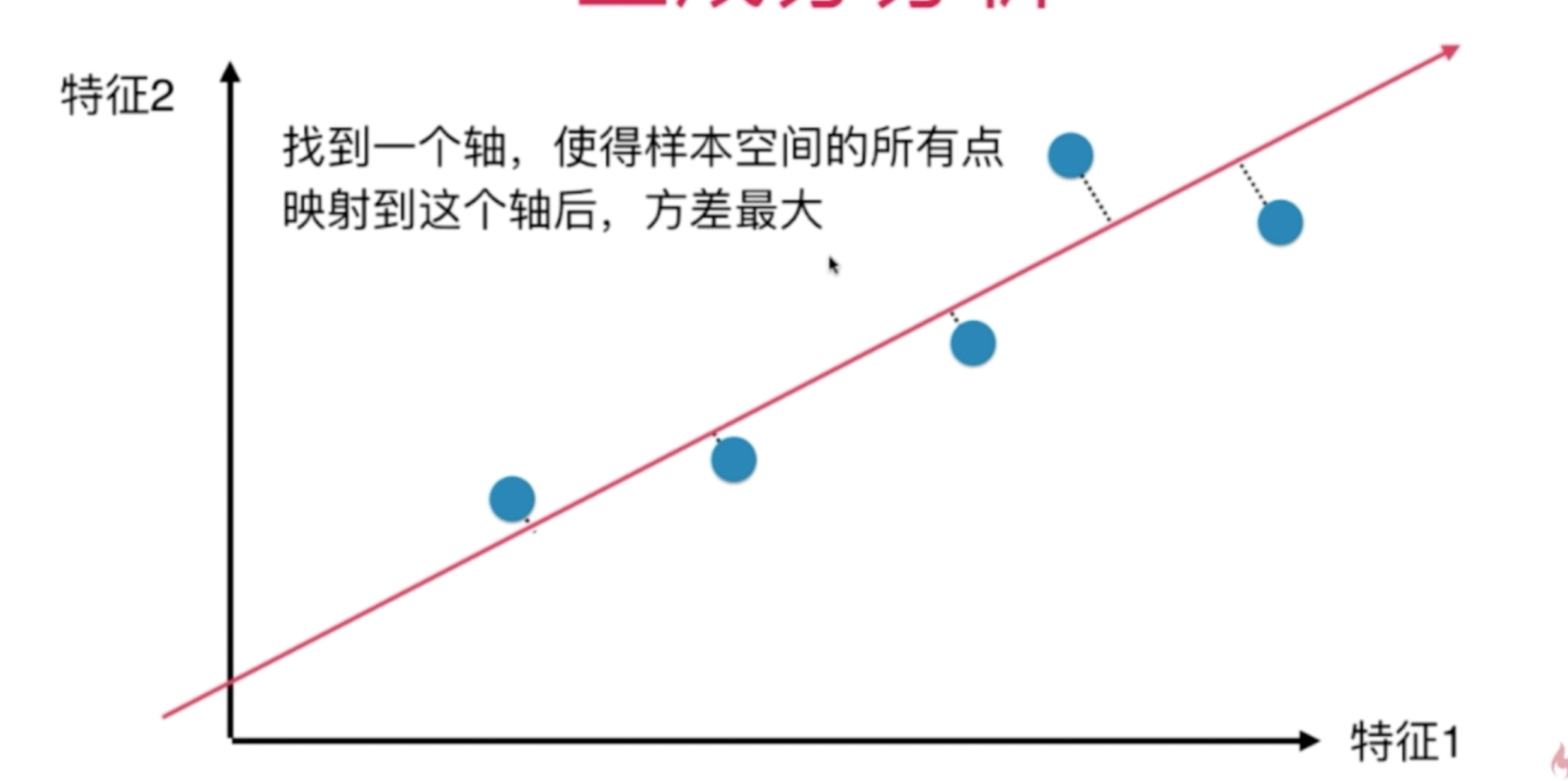


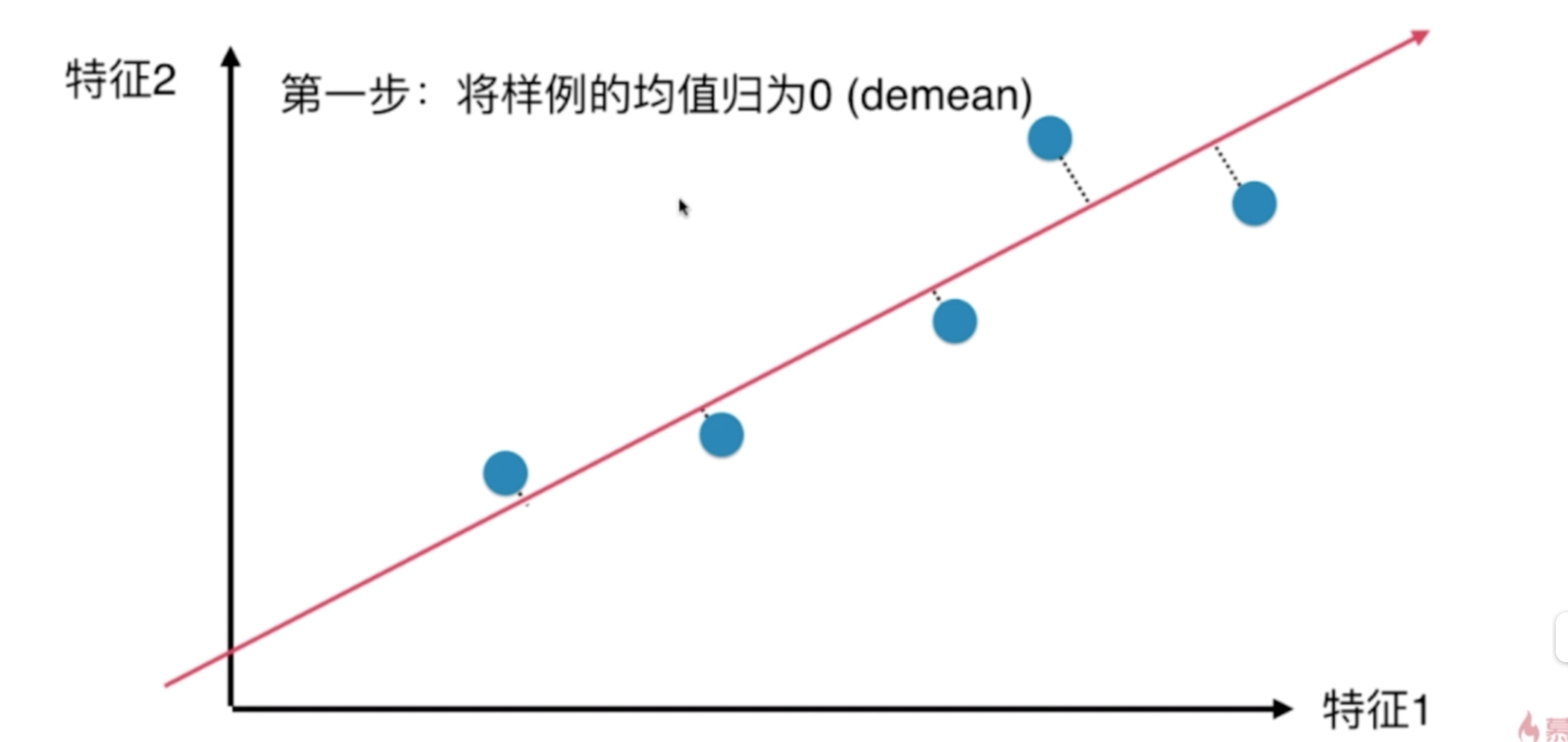
用这个斜线作为降维后的唯一一个轴，那么各个样本映射到这个轴上的值，间距比前面两个降维方案更大，更加趋近于这些点原来的分布情况。

那么问题来了：





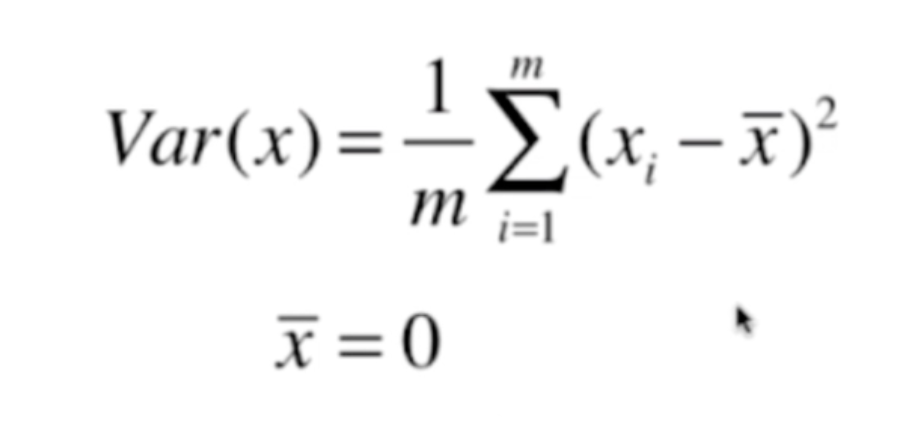




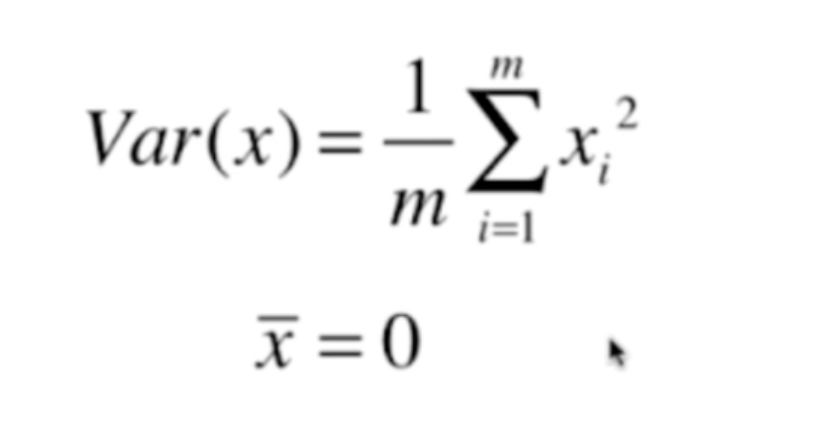
demean之后样本的分布就变成：



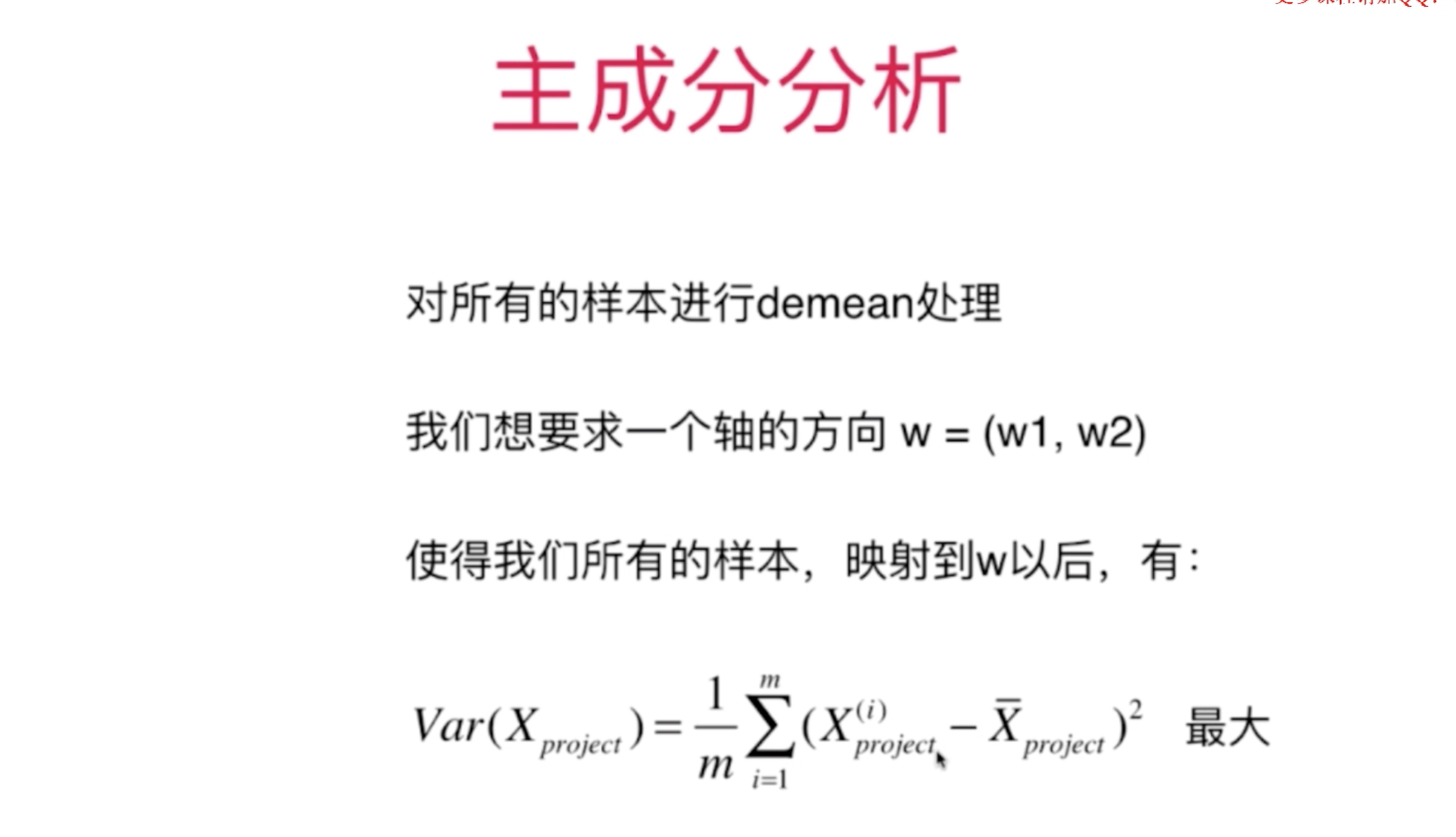
这样做是为了方便公式计算：



就可以转变为：



其中 Xi 是映射到新的轴后的样本值



其中 X(i)project 和 Xproject 的平均值 都是向量。那么换算到矩阵运算的公式就是：

